

Hoax Detection

By:  KEDATA

Hari Ke-1

<https://github.com/eppofahmi/hoaxdetection>


Definisi



KBBI Daring

Cari

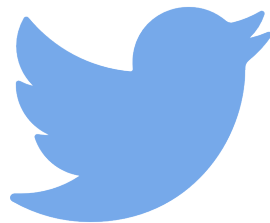
Seputar Laman

 Informasi: Temukan bantuan menggunakan KBBI Daring [di sini](#).



ho.aks

n cak berita bohong: *mereka mengumpulkan -- yang lalu lalang di banyak milis*



Program yang digunakan

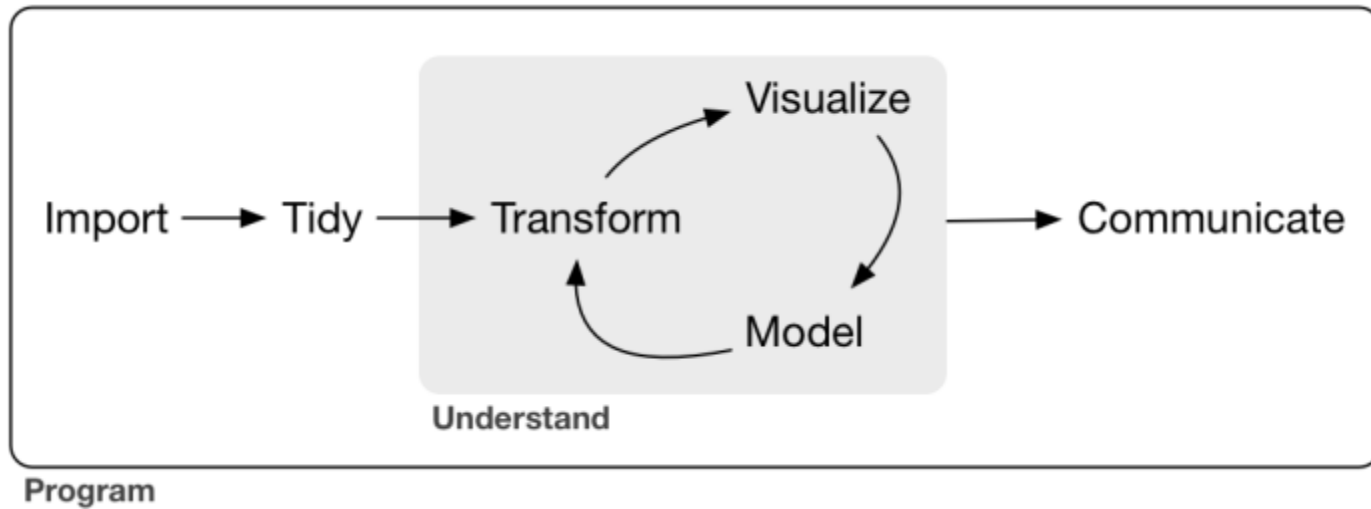


Beberapa alat yang akan digunakan dalam pelatihan ini juga sering digunakan oleh data scientist di seluruh dunia.

Rstudio digunakan untuk menulis dan menjalankan skrip **R**.
Github akan digunakan untuk membuat version control dan berbagai pengetahuan/masalah dengan orang lain.

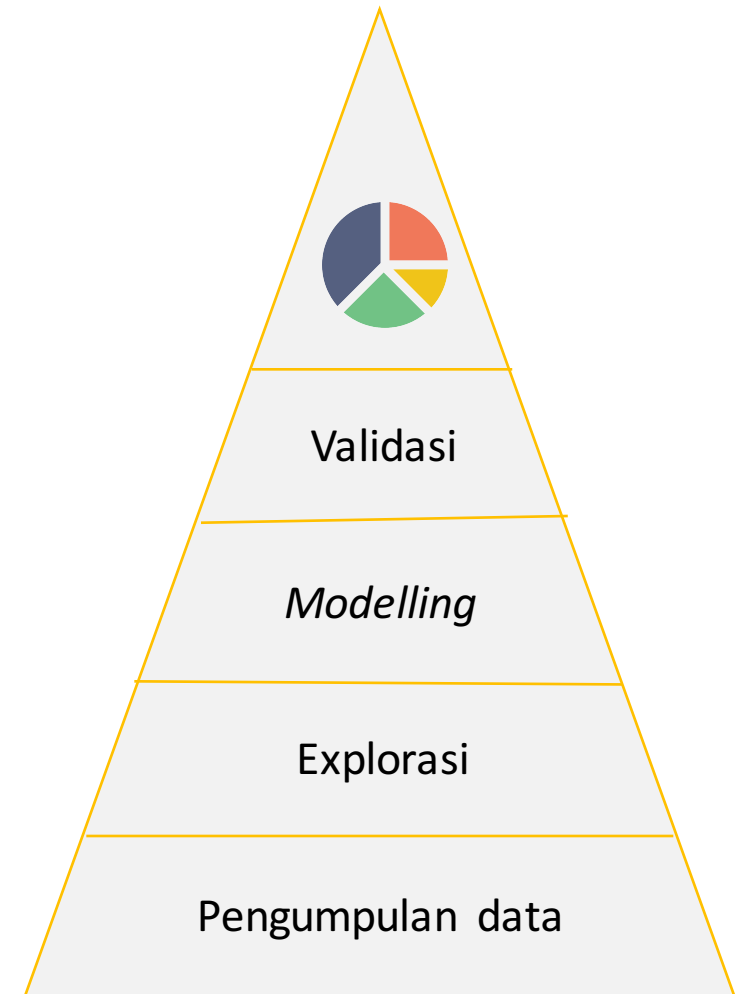


Cara berpikir data sains

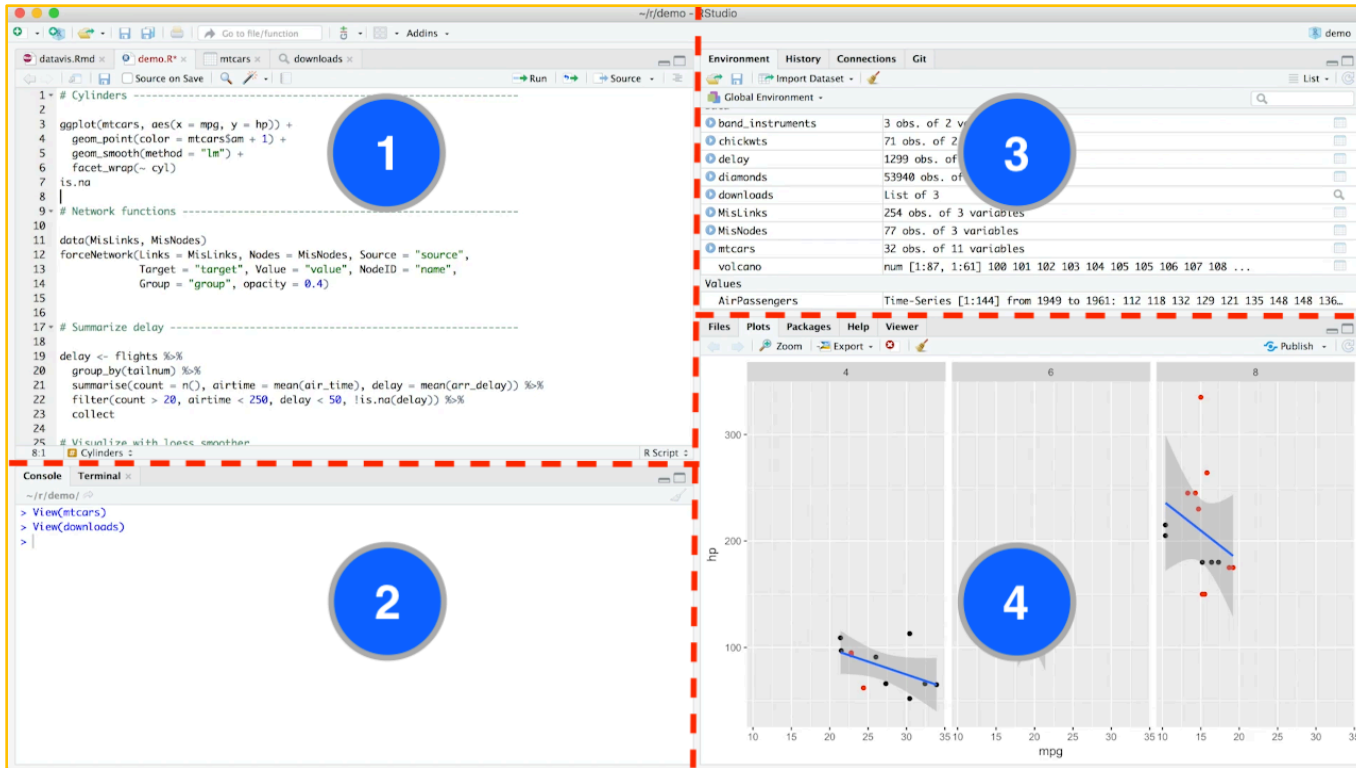


Data sains adalah bidang multi disiplin yang membutuhkan disiplin tinggi dalam mempelajarinya. Prosesnya bisa diawali dengan memastikan kita memiliki data dan bisa memasukan data pada program atau bahasa pemrograman yang akan digunakan.

Keterampilan yang disarankan untuk dikuasai selain domain knowledge yang jelas seorang *data scientist* juga perlu menguasai metode pengolahan data, membuat model, dan mengomunikasikannya.



Menggunakan R dan RStudio



1. Bagian pertama digunakan untuk menulis/membuat dan menjalankan perintah
2. Bagian kedua digunakan untuk melihat hasil dari skrip yang dijalankan atau untuk menulis skrip secara langsung
3. Bagian tiga digunakan untuk melihat objek yang menggunakan skrip, dan koneksi dengan lingkungan lain.
4. Bagian keempat digunakan untuk melihat hasil visualisasi, data yang terdapat directory yang digunakan, serta paket-paket yang terinstall, dll

Menulis skrip di R

1. Dasar
 - Membuat
 - Menulis
 - Menyimpan
 - Menggunakan R sebagai kalkulator
2. Library dan Package
 - Memilih
 - Menginstall
 - Memanggil
 - Menggunakan fungsi dari Package
3. Memahami fungsi dari library
 - Fungsi summary
 - Fungsi str
 - Fungsi class
 - Fungsi paste

nama_fungsi(argument)

Code	Dibaca
#	Comment
= atau <-	Equals to
%>%	Pipe or then
"..."	The object is string
[... , ...]	Rows , column
\$ (e.g. tweets\$retweet)	Part of (e.g. read column retweet from data with name tweet)
str(...)	Data structure
?help	Get some help from documentation

Mengenal Jenis Data

1. Jenis Data
 - Vectors
 - Lists
 - Data Frame
2. Ekspor Impor Data
 - Impor data jenis data csv, xlsx, txt
 - Ekspor data xlsx, csv, dan txt

Nama_kolom1	Nama_kolom2
Nama_a	text_1
Nama_b	text_2
Nama_c	text_3
Nama_d	text_4
Nama_e	text_5
Nama_f	text_6
Nama_g	text_7
Nama_f	text_8

Pre-processing Data

1. **Memilih kolom**
2. **Memfilter baris**
3. **Merename nama kolom**
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(lowercasing)
10. Tokenisasi

[illegible]

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. **Memisah dan menggabungkan**
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(lowercasing)
10. Tokenisasi

[illegible]

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. **Memisah dan menggabungkan**
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(*lowercasing*)
10. Tokenisasi

[illegible]

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. **Membuat kolom baru**
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(lowercasing)
10. Tokenisasi

[illegible]

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. **Menggabungkan data**
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(lowercasing)
10. Tokenisasi

[illegible][illegible]

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
- 6. Menggabungkan data**
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan (*lowercasing*)

ID				
			NA	NA
			NA	NA

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. Menggabungkan data
- 7. Menghilangkan url**
8. Menghilangkan username
9. Menyeragamkan tulisan
(*lowercasing*)
10. Tokenisasi

"#IwanFals jadi tamu kehormatan Menteri Desa, Marwan Jafar I
<http://www.kemendesa.go.id/berita/1421/menteri-marwan-undang-iwan-fals-sebagai-tamu-kehormatan-rakornas-desa> ..."



"#IwanFals jadi tamu kehormatan Menteri Desa, Marwan Jafar I
..."

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
- 8. Menghilangkan username**
9. Menyeragamkan tulisan
(*lowercasing*)
10. Tokenisasi

@KholiliLili @ansor_jatim @marwan_jafar @jarkom_desa
sabar, Selasa (31/3) akan dilaunching



sabar, Selasa (31/3) akan dilaunching

Pre-processing Data

1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
- 9. Menyeragamkan tulisan (*lowercasing*)**
10. Tokenisasi

"#IwanFals jadi tamu kehormatan Menteri Desa, Marwan Jafar I
<http://www.kemendesa.go.id/berita/1421/menteri-marwan-undang-iwan-fals-sebagai-tamu-kehormatan-rakornas-desa> ..."



"#IwanFals jadi tamu kehormatan Menteri Desa, Marwan Jafar I ..."



"jadi tamu kehormatan menteri desa, marwan jafar"

Pre-processing Data

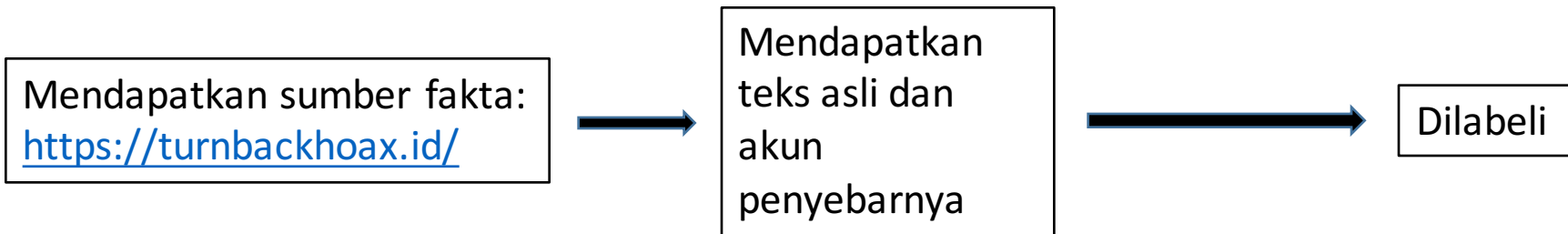
1. Memilih kolom
2. Memfilter baris
3. Merename nama kolom
4. Memisah dan menggabungkan
5. Membuat kolom baru
6. Menggabungkan data
7. Menghilangkan url
8. Menghilangkan username
9. Menyeragamkan tulisan
(*lowercasing*)

10. Tokenisasi

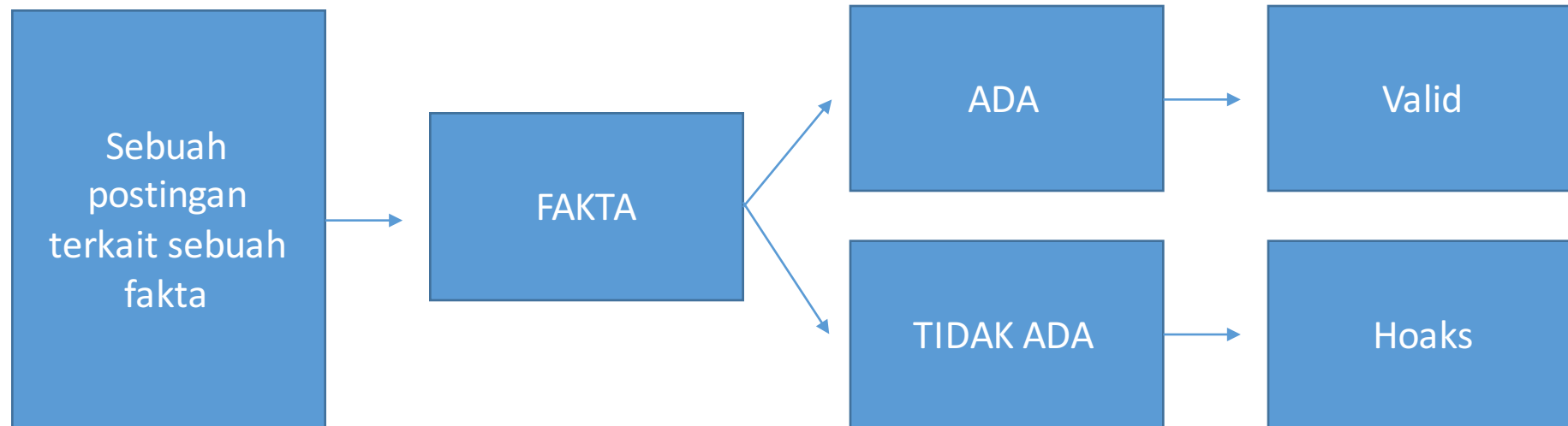
Text	Token	Jumlah
aku adalah raja yang dipilih raja	aku	1
aku adalah raja yang dipilih raja	adalah	1
aku adalah raja yang dipilih raja	raja	2
aku adalah raja yang dipilih raja	yang	1
aku adalah raja yang dipilih raja	dipilih	1
aku adalah raja yang dipilih raja	raja	2

Intro to machine learning

Membuat Data Latih



CARA KERJA MEMBERI LABEL



Persiapan hari ke-2

1. Membuat akun github: <https://github.com/>
2. Membuat/menyiapkan akun twitter

Hari Ke 2

Melabeli Data

No	Username	Follower	Text	Category	Label_1	Label_2	Label_3	Final
1								
2								
3								

Category:

1. Politik
2. Sosial:
3. Kesehatan:
4. Bisnis

Pengambilan Data

Syarat:

1. Memiliki akun twitter
2. Membuat API:

<https://developer.twitter.com/en/apps>